Relaxed uncertainty relations and information processing

Greg Ver Steeg* Stephanie Wehner†
Institute for Quantum Information, California Institute of Technology, Pasadena CA 91125, USA

May 5, 2009

Abstract

We consider a range of "theories" that violate the uncertainty relation for anti-commuting observables derived in [JMP, 49, 062105 (2008)]. We first show that Tsirelson's bound for the CHSH inequality can be derived from this uncertainty relation, and that relaxing this relation allows for non-local correlations that are stronger than what can be obtained in quantum mechanics. We continue to construct a hierarchy of related non-signaling theories, and show that on one hand they admit superstrong random access encodings and exponential savings for a particular communication problem, while on the other hand it becomes much harder in these theories to learn a state. We show that the existence of these effects stems from the absence of certain constraints on the expectation values of *commuting* measurements from our non-signaling theories that are present in quantum theory.

1 Introduction

In any physical theory, we may consider measurements M that when applied to a state ρ result in some measurement outcome k with probability $P(k|M)$, depending on ρ. A crucial element in characterizing the power of any physical theory lies in understanding what probability distributions are indeed possible. Quantum theory, for example, imposes strict limits on such distributions, which greatly affects our ability to perform information processing tasks [39]. One of these limitations is commonly known as an *uncertainty relation*. We may for example ask whether for some fixed choice of measurements M_1 and M_2 there even exists any state such that both distributions can be arbitrarily well defined. That is, is it possible that there exist outcomes k_1 and k_2 such that $P(k_1|M_1) = P(k_2|M_2) = 1$? Curiously, it turns out that in quantum theory there do indeed exist pairs of measurements M_1 and M_2 for which this is impossible. Another limitation is known as the *strength of non-local correlations*, which are restrictions on the joint probability distributions we can obtain when performing measurements on spatially separated systems. Classically, these limitations are known as Bell inequalities, and the corresponding limitations in the quantum case are referred to as Tsirelson bounds.

*gregv@caltech.edu
†wehner@caltech.edu

Since quantum mechanics imposes very stringent restrictions on the possible distributions [22], we would much like to understand their extent and implications. To this end, it is instructive to remove some of these restrictions and investigate how our ability to perform information processing tasks changes as a result. In this work, we will relax an uncertainty relation, which greatly affects our ability to solve communication and coding tasks. We will also see that the different kinds of restrictions are very closely related and show that for example Tsirelson's bound for the CHSH inequality is a consequence of the uncertainty relation of [42].

1.1 Previous work

Previous work has focused on investigating one particular restriction imposed by quantum mechanics, namely its limits on non-local correlations. Indeed, the existence of non-local correlations in quantum mechanics that are stronger than those allowed by local realism [9], but yet strictly weaker than those consistent with the no-signaling principle [31] poses an enigma to the understanding of the foundations of quantum physics. What are the properties of quantum mechanics that disallow these stronger correlations [24]? And, what possibilities would be opened by the existence of these correlations? Much of the work exploring these questions has focused on the "box paradigm" that was initially inspired by the CHSH inequality [16]. This particular Bell inequality [9] can be cast into a form of a simple game between two players, Alice and Bob. When the game starts, Alice and Bob are presented with randomly and independently chosen questions $s \in \{0,1\}$ and $t \in \{0,1\}$ respectively. They win if and only if they manage to return answers $a \in \{0,1\}$ and $b \in \{0,1\}$ such that $s \cdot t = a \oplus b$. Alice and Bob may thereby agree on any strategy before the game starts, but may not communicate afterwards. Classically, that is in any model based on local realism, this strategy consists of shared randomness. It has been shown [16] that for any such strategy we have

$$\gamma := \frac{1}{4} \sum_{s,t \in \{0,1\}} \Pr[s \cdot t = a_s \oplus b_t] \leq \frac{3}{4},$$

where $\Pr[s \cdot t = a_s \oplus b_t]$ is the probability that Alice and Bob return winning answers a_s and b_t when presented with questions s and t. Quantumly, Alice and Bob may choose any shared quantum state together with local measurements as part of their strategy. This allows them to violate the inequality above, but curiously only up to a value

$$\gamma \leq \frac{1}{2} + \frac{1}{2\sqrt{2}},$$

known as Tsirelson's bound [14, 15]. We will see later that there exists a state $|\Psi\rangle_{AB}$ shared by Alice and Bob that achieves this bound when Alice and Bob perform measurements given by the observables $A_0 = B_0 = X$ and $A_1 = B_1 = Z$ where we use A_s and B_t to denote the measurement corresponding to questions s and t respectively. The non-signaling principle that disallows faster than light communication between Alice and Bob alone does not impose such a restrictive bound. Hence, Popescu and Rohrlich [31, 32, 33] raised the question why nature is not more 'non-local'? That is, why does quantum mechanics not allow for a stronger violation of the CHSH inequality up to the maximal value of 1? To gain more insight into this question, they constructed a toy-theory based on so-called PR-boxes [45]. Each such box takes inputs $s, t \in \{0,1\}$ from Alice and Bob respectively and simply outputs randomly chosen measurement outcomes a_s,b_t such that $s \cdot t = a_s \oplus b_t$. Each such box can be used exactly once, and no notion of post-measurement states

exists. Note that Alice and Bob still cannot use this box to transmit any information. However, since we have for all s and t that $\Pr[s \cdot t = a_s \oplus b_t] = 1$, Tsirelson's bound is clearly is violated. It is interesting to consider how our ability to perform information processing tasks changes, if PR-boxes indeed existed. For example, it has been shown that Alice and Bob can use such PR-boxes to compute any Boolean function $f : \{0,1\}^{2n} \to \{0,1\}$ of their individual inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ by communicating only a *single* bit [39], which is even true when the boxes have slight imperfections [11].

Much interest has since been devoted to the study of such PR-boxes and their generalizations known as non-local boxes [23, 13, 19, 28, 5, 6, 27]. In particular, they have been incorporated in a very nice way into generalized non-signaling theories (GNST) due to Barrett [8] (the relation of such theories to generalizations of quantum theory is due to Hardy [24]) as a means of exploring foundational questions in quantum information. Intuitively, such theories allow for "boxes" involving many more inputs for one or more players/systems, and also allow for some transformations between such boxes. Both theories seek out physically motivated properties that single out quantum mechanics from other theories such as the classical world. These theories have also found interesting applications in deriving new bounds for quantum mechanics itself, e.g., monogamy of entanglement [37].

In such a theory, n-partite states are characterized by the probabilities of obtaining certain outcomes when performing a fixed set of local fiducial measurements on each system. For example, to describe a non-local box, consider a bipartite system, where Alice holds the first and Bob the second system. We will label both Alice and Bob's measurements using X and Z in analogy to the quantum setting. For convenience we will also label the outcomes using $a, b \in \{0,1\}$, where the actual outcomes of X and Z in the quantum setting could be recovered as $(-1)^a$, and use $p(A|M)$ to denote the probability of obtaining outcomes A for measurements M. A non-local box is now given by the probabilities $p(0,0|X,X) = p(0,0|X,Z) = p(0,0|Z,X) = 1/2$, $p(1,1|X,X) = p(1,1|X,Z) = p(1,1|Z,X) = 1/2$, $p(0,1|Z,Z) = p(1,0|Z,Z) = 1/2$ and $p(A|M) = 0$ otherwise. We will describe such theories in more detail in section 4. We will also refer to GNST using the commonly used term "box-world".

1.2 Relaxed uncertainty relations

Even when allowing more than two measurements and outcomes, such boxes remain very artificial constructs and it is not quite clear how they relate to quantum theory. In this note, we hope to provide a more intuitive understanding by showing that superstrong correlations can indeed be obtained by relaxing an uncertainty relation known to hold in quantum theory. Consider *any* anti-commuting observables $\Gamma_1, \ldots, \Gamma_{2n}$ satisfying

$$\{\Gamma_j, \Gamma_k\} = 0$$

whenever $j \neq k$ and

$$\Gamma_j^2 = \mathbb{I},$$

for any $j \in [2n]$, and let $\Gamma_0 = i\Gamma_1 \ldots \Gamma_{2n}$ (see section 2 on how to construct such operators). It was shown in [42] that any quantum state obeys

$$\sum_{j=0}^{2n} \mathrm{Tr}\left(\Gamma_j \rho\right)^2 \leq 1, \tag{1}$$

which also lead to several entropic uncertainty relations for such observables. To see why Eq. (1) itself can be understood as an uncertainty relation note that $\mathrm{Tr}(\Gamma_j\rho)$ is the expectation value of measuring the observable Γ_j on ρ. The probability of obtaining a measurement outcome $b \in \{\pm 1\}$ can furthermore be written as $p(b|\Gamma_j) = 1/2 + b\,\mathrm{Tr}(\Gamma_j\rho)/2$. Hence, $\mathrm{Tr}(\Gamma_j\rho)$ can also be understood as the bias towards a particular measurement outcome. Eq. (1) now tells us that this bias cannot be arbitrarily large for all measurements Γ_j. Note that we could rewrite the condition of Eq. (1) as $||v||_2^2 \leq 1$ where $v = (\mathrm{Tr}(\Gamma_1\rho), \ldots, \mathrm{Tr}(\Gamma_{2n}\rho))$. Whereas the uncertainty relations of [42] may appear unrelated to the problem of determining the strength of non-local correlations, we will see later that Tsirelson's bound for the CHSH inequality is in fact a consequence of Eq. (1), when we use the fact that local anti-commutation and maximal violations of the CHSH inequality are closely related [14, 38, 34]. Thus, as one might intuitively guess, bounds for the strength of non-local correlations are indeed closely related to uncertainty relations, and such connections have been observed in a different form by [28, 8].

What happens if we merely ask for $||v||_p^p \leq 1$, where $||\cdot||_p$ is the p-norm of the vector v? Since Eq. (1) must hold for any quantum state, that is for any positive semi-definite matrix ρ with $\mathrm{Tr}(\rho) = 1$, it is clear that this allows operators ρ which are no longer positive semi-definite. In the spirit of Barrett's GNST, we will however restrict ourselves to allowing a particular set of fiducial measurements only, for which the probabilities will remain positive and thus well-defined. In section 3, we will describe a hierarchy of such "theories" in detail, and investigate their power with respect to non-local correlations and information processing problems. In particular, we will see that

- For the CHSH inequality, we can obtain at most

$$\gamma = \frac{1}{2} + \frac{1}{2(2)^{1/p}} \text{ for } ||v||_p^p \leq 1.$$

  where in the limit of $p \to \infty$ the right-hand side becomes 1, and we have a state that acts analogous to a non-local box.

- Furthermore, any unique XOR-game can be played with perfect success for $p \to \infty$.

It is instructive to consider what our relaxed uncertainty relation means in the case of a single qubit. Note that for quantum mechanics we have $p = 2$ in which case Eq. (1) corresponds to the statement that v must lie inside the Bloch sphere. Allowing different values of p now constraints us to the corresponding p-spheres as depicted in Figure 1.2. It is interesting to consider that even though for $p > 2$ we obtain non-local correlations that are *stronger* than what quantum theory allows, we now have a *weaker* uncertainty relation than in quantum theory. It has previously been noted by Barrett [8] that GNST has no uncertainty relations for particular measurements. Our work makes this relation very intuitive. In particular, for the case of $p \to \infty$ corresponding to a non-local box we essentially place no restrictions on the bias $\mathrm{Tr}(\Gamma_j\rho)$ at all. Since Eq. (1) leads to the entropic uncertainty relations on which the security of the protocols in the bounded-quantum-storage model [17, 18, 43] is based, it may be worth considering how certain cryptographic tasks change in the setting of non-local boxes. Indeed, it has recently been shown [44] that privacy amplification fails in a world based on non-local boxes. Whereas it is known that cryptographic tasks such as bit commitment and oblivious transfer are compatible with the no-signaling principle [13], little is known about them in general theories [7].

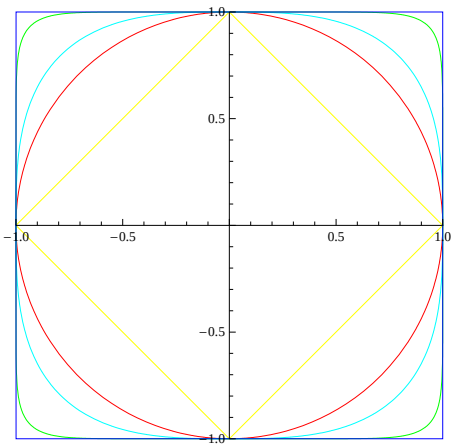

Figure 1: p-norm unit circles in dimension 2 for $p = 1, 2, 3, 10, 10000$

It should be noted that except for a single qubit, Eq. (1) is of course only a necessary and not a sufficient condition for $\rho \geq 0$. In higher dimensions, such relations are much more involved, but have been obtained for certain operators [26, 10, 21] and also some operators relating more closely to unbiased measurements [41]. Relaxing this particular uncertainty relation is thus only one way to go. Yet, due to the rich structure of the Clifford algebra of operators $\Gamma_1, \ldots, \Gamma_{2n}$ and their central importance for entropic uncertainty relations and so-called XOR non-local games (also known as two-party correlation inequalities) with 2 measurement outcomes, this small relaxation allows us to gain some insights into their role in quantum information processing tasks.

1.3 Information processing in generalized non-local theories

Inspired by these relaxations in terms of an operator ρ, we then construct a hierarchy of p-GNST theories exhibiting similar constraints. For such theories, we identify a single gbit (defined in [8]) with a single qubit obeying the relaxed uncertainty relations above. That is, we will think of a single gbit as allowing three fiducial measurements labeled X, Z and Y in analogy to the quantum case. Whereas this choice is of course again quite arbitrary, and heavily inspired by the quantum setting, it will allow us to gain a slightly better understanding of the relation of "box-world" and quantum theory later on. We show that the states we allow above, as well as states in p-GNST's have several properties that set them apart from quantum theory. In particular, we will see that

- In p-GNST, there exists superstrong random access encodings. For example, there exists an encoding of $N = 3^n$ bits into $(2n+1)^{3/p}n$ *gbits* such that we can retrieve any bit with probability $1-\varepsilon$ for $\varepsilon = 2\exp(-(2n+1)^{1/p}/2)$. Quantumly on the other hand it is known that we require at least $(1 - h(1-\varepsilon))N$ *qubits* to encode N classical bits with the same recovery probability, where h denotes the binary Shannon entropy.

- As a consequence, in p-GNST there exist single server PIR scheme with $O(\text{polylog}(N))$ bits of communication for an N bit database with large N, whereas quantumly $\Omega(N)$ bits are needed.

- On the other hand, we show that in GNST it becomes much harder to learn a state in the sense of [1]. In fact, unlike in the quantum setting, we can essentially not ignore even a small part of the information we are given about a state.

Note that we thereby compare *units of information*, gbits vs. qubits, irrespective of a physical dimension, where gbits were previously defined in [8]. It may not be surprising that such effects exist for Hermitian operators ρ, when all we essentially demand is that the condition $||v||_p^p \leq 1$ is obeyed for any set of anti-commuting measurements. However, it will be interesting to consider why for example the superstrong random access code encodings we find above are disallowed in quantum theory, but allowed in GNST.

1.4 Commuting measurements

Although the results of local measurements suffice to describe quantum states [24], our results suggest that building a toy-theory around local measurements acting on fixed systems alone (such as GNST) may miss part of the flavor when considering some applications. Quantum mechanics has a rich structure of commuting and anti-commuting measurements built in which make no particular reference to locality. Uncertainty relations impose restrictions for non-commuting measurements, such as for example the anti-commuting measurements $\Gamma_1, \ldots, \Gamma_{2n}$. However, we will see in section 2.4 that also certain sets of commuting measurements cannot have arbitrary expectation values when measured on a particular state ρ. As a simple example, consider a 2 qubit system shared between Alice and Bob, and consider the measurement $X \otimes \mathbb{I}$, $\mathbb{I} \otimes X$ and $X \otimes X$. Suppose that we have $\mathrm{Tr}((X \otimes \mathbb{I})\rho) = \mathrm{Tr}((\mathbb{I} \otimes X)\rho) = 1$. This tells us that when Alice and Bob measure X locally, they obtain an outcome of '1' each with probability 1. However, the measurement of $X \otimes X$ can very intuitively be viewed as Alice and Bob performing a local measurement of X and taking the product of their outcomes. Hence, we do not expect a simultaneous assignment of $\mathrm{Tr}((X \otimes X)\rho) = -1$ to be consistent with the previous two expectation values. We will formalize this intuition in section 2.4, where we will derive a series of conditions such expectation values must obey which in spirit is similar to [22].

GNST does satisfy these conditions for measurements that commute because they act on different subsystems. It does not exhibit any inconsistencies otherwise, as no commutation relations are defined for measurements on the same system. The issue of such inconsistencies is further circumvented by the simple fact that a non-local box can only be used once, and there is no notion of subsequent measurements on the same system. This of course is perfectly adequate for studying the strength of non-local correlation between two space-like separated systems for example, and led to such perplexing results as [39]. We will however see that it is essentially this lack of additional constraints that allows us to form superstrong random access codes for example, and may indicate that using "box-world" to investigate the role of the strength of non-local correlations within quantum theory itself is possibly doomed to fail. It also indicates why defining a consistent notion of 'post-measurement' states for non-local boxes is quite difficult, since many constraints that would allow such a task to succeed are simply not present in box-world.

To see how box-world differs from quantum theory consider the measurements $M_1 = X \otimes Z$, $M_2 = Z \otimes X$ and $M_3 = -XZ \otimes XZ$. These are related in exactly the same way as the measurements we considered above, except that in GNST there is no notion that M_1 and M_2 commute. Yet, we intuitively expect similar conditions to hold as for the measurements above when trying to form an analogy to the quantum setting. Indeed, one can easily construct a unitary transformation that

maps the measurements M_1, M_2 and M_3 into a form analogous to the above, where two of the measurements act on different systems. [1] In GNST, however, the separation into different systems is always a given, which may lead to difficulties when examining some problems which are not really concerned with correlations among two distant systems alone, but to information processing in general.

1.5 Outline

Whereas we only examine a very small piece of the puzzle, our work hopes to shed some light on the relation between uncertainty relations, non-local correlations and the role of above mentioned consistency constraints in information processing. In section 2 we first explain the basic concepts we need to refer to. commuting measurements in more detail. In section 3 we then define a range of simple "theories" obtained by relaxing the uncertainty relation for anti-commuting observables. To highlight the analogy with non-local boxes, we then define a range of similar GNST-like theories in section 4. In sections 5, 6, and 7 we then investigate the power of such theories with respect to non-local correlations, random access codes, and information processing problems respectively. In section 2.4 we then investigate why such effects are possible within GNST, but not in quantum theory. Table 7.4 summarizes similarities and differences among theories.

2 Preliminaries

2.1 Basic concepts

In the following, we write $[n] := \{1, \ldots, n\}$ and use X, Z and Y to denote the well-known Pauli matrices [29]. We also speak of a *string of Paulis* to refer to a matrix of the form

$$S_{ab} := X^{a_1} Z^{b_1} \otimes \ldots \otimes X^{a_n} Z^{b_n}, \tag{2}$$

with $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$ and $a_j, b_j \in \{0, 1\}$. We sometimes write the Pauli operator acting on subsystem j, with identity on the other subsystems as

$$X_j = \mathbb{I}^{\otimes j-1} \otimes X \otimes \mathbb{I}^{\otimes n-j-1}$$

The *Pauli basis expansion* of a density matrix ρ is given by $\rho = (\mathbb{I} + \sum_{a,b} s_{ab} S_{ab})/d$, where we call s_{ab} the *coefficient* of S_{ab}. Consider the form $f(a, b, a', b') = (a, b') + (a', b)$, where we write $(a, b) = \sum_j a_j b_j \mod 2$. It it straightforward to convince yourself that for any pair S_{ab} and $S_{a'b'}$ either $[S_{ab}, S_{a'b'}] = 0$ if $f(a, b, a', b') = 0$ or $\{S_{ab}, S_{a'b'}\} = 0$ if $f(a, b, a', b') = 1$. Whereas Eq. (1) holds for any choice of anti-commuting measurements, it is worth noting that in dimension $d = 2^n$ we can find at most $2n + 1$ anti-commuting operators given by

$$\begin{aligned} \Gamma_{2j-1} &= Y^{\otimes(j-1)} \otimes X \otimes \mathbb{I}^{\otimes(n-j)} \\ \Gamma_{2j} &= Y^{\otimes(j-1)} \otimes Z \otimes \mathbb{I}^{\otimes(n-j)}, \end{aligned}$$

for $j = 1, \ldots, n$ and $\Gamma_0 = i\Gamma_1 \ldots \Gamma_{2n}$. Note that for $n = 1$ we have $\Gamma_1 = X$, $\Gamma_2 = Z$, $\Gamma_0 = Y$ and Eq. (1) is equivalent to the Bloch sphere condition. We will also need the notion of a p-norm of a

[1] Consider $U = (\mathbb{I} \otimes H)\text{CNOT}(\mathbb{I} \otimes H)$

vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ which is defined as

$$||v||_p := \left(\sum_{j=1}^{n} |v_j|^p \right)^{1/p}.$$

Note that for $p = 2$ this is just the Euclidean norm. Of particular interest to us will also be the ∞-norm defined as $||v||_\infty := \lim_{p\to\infty} ||v||_p$ which can also be written as

$$||v||_\infty = \max(|v_1|, \ldots, |v_n|).$$

2.2 Probability distributions

Unlike previous descriptions of general probabilistic theories, our notation must be versatile enough to accommodate arbitrary choices of simultaneous commuting measurements, even if they do not act on separate subsystems. In quantum mechanics we may choose to measure $X \otimes X$ along with either $X \otimes \mathbb{I}, \mathbb{I} \otimes X$, or $Z \otimes Z, XZ \otimes XZ$. We will see that including this flexibility in a more general theory leads to new constraints.

First, we want to consider some finite set of measurements $\mathcal{O} = \{M_1, \ldots, M_N\}$ where without loss of generality we assume that each measurement has the same finite set of outcomes $\mathcal{A}$ and the $\mathcal{O}$ is ordered lexiocraphically. Although we initially impose no structure on $\mathcal{O}$, in analogy to quantum mechanics we consider certain collections of measurements $C \subseteq \mathcal{O}$ to have some property which directly corresponds to simultaneous measurability. In particular, we will consider the set of possible experiments

$$\mathcal{E} := \{C \subseteq \mathcal{O} \ \wedge \ \forall M_i, M_j \in C \ \mathrm{sim}(M_i, M_j) = 0\},$$

where "sim" is a predicate indicating simultaneous measurability that remains to be specified. Of particular concern to us will be the probability distributions p over the outcomes $A \in \mathcal{A}^{\times|C|}$ of some set of simultaneously performed measurements $C \in \mathcal{E}$. We use $p(A|C)$ to denote the probability of obtaining outcomes $A = (A_1, A_2, \ldots, A_{|C|}) \in \mathcal{A}^{\times|C|}$ for measurements $C \subseteq \mathcal{O}$ where we wlog take C to be ordered lexicographically. For simplicity, we will also write $p(A_1, \ldots, A_n|M_1, \ldots, M_n) := p((A_1, \ldots, A_n)|\{M_1, \ldots, M_n\})$.

What conditions do the functions $p : \mathcal{A}^{\times|C|} \times C \to [0, 1]$ have to fulfill be a valid probability distribution for any experiment $C \in \mathcal{E}$? We require that the following conditions need to be satisfied for *any* probability distribution

(1) Normalization: $\forall C \in \mathcal{E}, \sum_{A\in\mathcal{A}^{\times|C|}} p(A|C) = 1$.

(2) Positivity: $\forall C \in \mathcal{E}, \forall A \in \mathcal{A}^{\times|C|}, p(A|C) \geq 0$.

The next condition may appear unfamiliar at first glance. Intuitively it says that the distributions of outcomes we obtain for commuting measurements are independent of what other commuting measurements we perform.

(3) Independence:

$$\forall C, C' \in \mathcal{E} \text{ with } C \subseteq C', \quad p((A_1, \ldots, A_{|C|})|C) = \sum_{A_{|C|+1},\ldots,A_{|C'|}\in\mathcal{A}^{\times|C'|}} p((A_1, \ldots, A_{|C'|})|C'),$$

where, without loss of generality, we take the first $|C|$ outcomes to be associated with the measurements in C.

Throughout this text, we explore the result of choosing two different ways of choosing simultaneous measurements. First, we consider simultaneous measurements on distinct systems as reflected in the construction of non-local boxes. Second, we consider a more general notion of such measurements based on commutation relations as in quantum mechanics. Note that in the quantum case such sets of mutually commuting measurements induce a partitioning of the Hilbert space into different systems in the finite-dimensional setting [36, 22].

Consider the set of measurements $\mathcal{O}_P$ to be strings of Paulis on n-partite systems as defined in section 2.1. The two different notions of simultaneous measurements can now be expressed in two different choices of $\mathrm{sim}(M_i, M_j)$, leading to two different sets of realizable experiments. To capture the first notion, we let

$$\mathcal{E}_L := \{C \subseteq \mathcal{O}_P \ \wedge \ \forall M_i, M_j \in C \ \mathrm{local}(M_i, M_j) = 0\},$$

where $\mathrm{local}(M_i, M_j) = 0$ if and only if M_i and M_j act on different subsystems. For example, we have $\mathrm{local}(X \otimes \mathbb{I}, \mathbb{I} \otimes Z) = 0$. Second, we let

$$\mathcal{E}_C := \{C \subseteq \mathcal{O}_P \ \wedge \ \forall M_i, M_j \in C \ [M_i, M_j] = 0\},$$

where all commuting measurements are simultaneously observable, as in quantum mechanics. Clearly, $\mathcal{E}_L \subseteq \mathcal{E}_C$, since two measurements acting on two different subsystems commute.

When we restrict ourselves to $\mathcal{E}_L$ we can express the independence condition from above in the more familiar form of no-signaling:

(3') No-signaling:

$$\forall C, C' \in \mathcal{E}_L \text{ with } C \subseteq C', \quad p((A_1, \ldots, A_{|C|})|C) = \sum_{A_{|C|+1}, \ldots, A_{|C'|} \in \mathcal{A}^{\times |C'|}} p((A_1, \ldots, A_{|C'|})|C').$$

Intuitively, the no-signaling condition just dictates that the marginal distribution of a particular subset of systems is *independent* of the measurement choices on a disjoint subset of systems. Therefore, we can simplify our description of marginals of no-signaling distributions to just $p(A \in \mathcal{A}^{\times |C|}|C') = p(A|C)$, where the measurement choices on other parties are arbitrary. We will later see that imposing only the special case of the no-signaling condition, versus the full independence condition of (3), makes a crucial difference in the power of the resulting theory with respect to encoding information.

Example 2.1. *Consider the set of local experiments for two parties with* $\mathcal{A} = \{-1, 1\}, \mathcal{O} = \{X_1, Z_1, X_2, Z_2\}$. *Let the probability distribution* $p(A|C)$ *be described by the following table.*

A					
$(1,1)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	
$(1,-1)$	0	0	0	$\frac{1}{2}$	
$(-1,1)$	0	0	0	$\frac{1}{2}$	
$(-1,-1)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	
	$\{X_1, X_2\}$	$\{X_1, Z_2\}$	$\{Z_1, X_2\}$	$\{Z_1, Z_2\}$	C

Clearly, we have positivity, and the sum over each measurement setting (column) is 1. *Finally, note that the marginal probability distribution for either party is constant,* $\forall C \in \mathcal{E}_L, \forall A_1 \in \mathcal{A}, \sum_{A_2 \in \mathcal{A}} p((A_1, A_2)|C) = \frac{1}{2}$, *therefore this distribution is no-signaling.*

2.3 Moments

Any finite, discrete probability distribution has a dual representation in terms of a finite number of moments [40]. We define the product of the outcomes $A = (A_1, \ldots, A_{|C|}) \in \mathcal{A}^{\times|C|}$ of a collection of measurements $C \in \mathcal{E}$ as $A^* = \prod_{i=1}^{|C|} A_i$. The moment for this measurement is defined as

$$m(C) := \sum_{A \in \mathcal{A}^{\times|C|}} p(A|C) A^*. \tag{3}$$

Note that for the identity measurement this means $m(\mathbb{I}) = 1$ because of normalization. Also, if you consider the moment for some subset of C, by the independence principle this definition gives a unique value which does not depend on the choice of other measurements made simultaneously.

Since we will only be concerned with measurements with two outcomes $\mathcal{A} = \{\pm 1\}$, we now restrict ourselves to this case for simplicity. For the measurement of a single observable $C = \{M_1\}$ with outcome $A_1 \in \mathcal{A}$, we can easily recover the probabilities from the moments as

$$p((A_1)|\{M_1\}) = \frac{1}{2}\left(1 + A_1 \; m(\{M_1\})\right). \tag{4}$$

In subsequent notation, we will drop the brackets within parentheses when it increases readability.

Note that we can recover the probability for a specific set of outcomes $\hat{A} \in \mathcal{A}^{\times|C|}$ and measurements $C \in \mathcal{E}$ from these moments. Without loss of generality, let $C = \{M_1, \ldots, M_n\}$.

$$\begin{aligned}
& \frac{1}{2^n} \sum_{C' \subseteq C} m(C') \prod_{i, M_i \in C'} \hat{A}_i \\
= \; & \frac{1}{2^n} \sum_{C' \subseteq C} \left(\sum_{A \in \mathcal{A}^{\times|C'|}} p(A|C') \prod_{i, M_i \in C'} A_i \right) \prod_{i, M_i \in C'} \hat{A}_i \\
= \; & \frac{1}{2^n} \sum_{A \in \mathcal{A}^{\times|C|}} p(A|C) \sum_{C' \subseteq C} \prod_{i, M_i \in C'} A_i \hat{A}_i
\end{aligned}$$

The second line simply uses the definition of $m(C')$ and the third line uses the independence principle to write $p(A|C')$ in terms of $p(A|C)$, allowing us to move the sum over C' inside. Now note that the sum over C' can be broken into n sums over whether or not $M_i \in C'$. For each M_i, if it is in C' we get a factor of $A_i \hat{A}_i$, otherwise a factor of 1.

$$= \frac{1}{2^n} \sum_{A \in \mathcal{A}^{\times|C|}} p(A|C) \prod_{i=1}^{n} (1 + A_i \hat{A}_i)$$

Because the outcomes can only be ± 1, the sum can give us only 0 or 2.

$$\begin{aligned} &= \frac{1}{2^n} \sum_{A \in \mathcal{A}^{\times |C|}} p(A|C) \prod_{i=1}^{n} 2\delta_{A_i, \hat{A}_i} \\ &= \frac{1}{2^n} \sum_{A \in \mathcal{A}^{\times |C|}} p(A|C)\ 2^n \delta_{A, \hat{A}} \\ &= p(\hat{A}|C) \end{aligned}$$

2.4 Consistency constraints

We are now ready to investigate the constraints that arise due to simultaneous measurement of commuting observables and that will play a crucial role in understanding the differences between quantum theory and p-GNST. Imagine two commuting measurements $[M_i, M_j] = 0$, and their product $M_k = M_i M_j$. In quantum mechanics the outcome of the measurement M_k is the same as the product of the outcomes of M_i and M_j, which can be verified by expanding M_k in terms of M_i and M_j and using the fact that they have a joint eigenbasis. What happens if we take this to be true in any theory? If we are only allowed to make local measurements, then this is a moot point. We can only get $X \otimes X$ by measuring $X \otimes \mathbb{I}$ and $\mathbb{I} \otimes X$ and multiplying the results.

But if we are allowed to make any combination of commuting measurements, this will impose some interesting conditions. For example, in the quantum case we may have $M_1 = X \otimes X$, $M_2 = Z \otimes Z$ and $M_3 = XZ \otimes XZ$. To see that this has consequences in terms of the moments, consider the simple example where $m(M_1) = 1$ and $m(M_2) = 1$, which means that we will deterministically observe outcomes $A(M_1) = A(M_2) = 1$. Hence, $m(M_3) = -1$ should intuitively not be compatible with these two moments for M_1 and M_2.

How can we formalize these conditions? For example, Eq. (3) gives us that

$$m(M_1 M_2) = m(M_1, M_2),$$

if we insist that outcomes of products of measurements equal the product of outcomes of individual measurements. For a given set of commuting measurements $C = \{M_1, \ldots, M_m\}$ with $M_j^2 = \mathbb{I}$, let $s(M)$ be the 2^m element vector whose k-th entry is given by

$$s := [s(C)]_k := M_1^{k_1} M_2^{k_2} \ldots M_m^{k_m}, \tag{5}$$

with $k \in \{0,1\}^m$ in lexicographic order. We now define the *moment matrix* K_s by letting the entry in the i-row and j-th column be given by

$$[K_s]_{ij} := m(s_i s_j)/2^m.$$

Claim 2.2 (Adapted from Wainwright and Jordan [40])**.** *Let $C = \{M_1, \ldots, M_m\}$ be a set of commuting measurements. Then $K_s \geq 0$ if and only if p is a probability distribution (satisfying constraints (1) and (2)).*

Proof. In addition to K_s, we define two more $2^m \times 2^m$ matrices, whose components are labeled by vectors $i, j \in \{0,1\}^m$ in lexicographic order as

$$\begin{aligned} [P]_{ij} &= \delta_{ij} p(A = ((-1)^{i_1}, \ldots, (-1)^{i_m})|C). \\ [B]_{ij} &= \frac{1}{2^{m/2}} (-1)^{i \cdot j}, \end{aligned}$$

It is easily verified that B is a unitary matrix. Note that B is an example of a Hadamard matrix. Now we will show that $K_s = BPB^\top$.

$$\begin{aligned}
\left[BPB^\top\right]_{ij} &= \frac{1}{2^m} \sum_{k,l\in\{0,1\}^m} (-1)^{i\cdot k}\ \delta_{kl}\ p(((-1)^{k_1},\ldots,(-1)^{k_m})|C)(-1)^{l\cdot j} \\
&= \frac{1}{2^m} \sum_{k\in\{0,1\}^m} (-1)^{k\cdot(i\oplus j)} p(((-1)^{k_1},\ldots,(-1)^{k_m})|C) \\
&= \frac{1}{2^m} \sum_{k\in\{0,1\}^m} \prod_{t=1}^{m} ((-1)^{k_t})^{(i_t\oplus j_t)} p(((-1)^{k_1},\ldots,(-1)^{k_m})|C) \\
&= \frac{1}{2^m} \sum_{A\in\mathcal{A}^{\times|C|}} \prod_{t=1}^{m} A_t^{i_t} A_t^{j_t} p(A|C) \\
&= \frac{1}{2^m} m(s_i s_j) = [K_s]_{ij}
\end{aligned}$$

Clearly, if the probabilities $p(A|C)$ are non-negative (2), then $P \geq 0$ if and only if $K \geq 0$ since B is unitary. Similarly, the fact that $m(\mathbb{I}) = 1$, B is unitary and the trace is cyclic ensures that p satisfies condition (1). □

Example 2.3. *As an example, consider the case of two commuting measurement M_1 and M_2 with $M_3 = M_1 M_2$. We have $s = (\mathbb{I}, M_1, M_2, M_3)$ and*

$$\mathbf{K}_s = \begin{pmatrix} m(\mathbb{I}) & m(M_1) & m(M_2) & m(M_1M_2) \\ m(M_1) & m(\mathbb{I}) & m(M_3) & m(M_2) \\ m(M_2) & m(M_3) & m(\mathbb{I}) & m(M_1) \\ m(M_3) & m(M_3) & m(M_1) & m(\mathbb{I}) \end{pmatrix} \equiv \begin{pmatrix} 1 & a & b & c \\ a & 1 & c & b \\ b & c & 1 & a \\ c & b & a & 1 \end{pmatrix}$$

Demanding that the eigenvalues of this matrix, $\lambda = ((1+a-b-c),(-1+a+b-c),(-1+a-b+c),(1+a+b+c))$, be non-negative is enough to ensure that $\mathbf{K}_s \succeq 0$. Using the Sylvester criteria, we get the alternate constraints that each moment $|a,b,c| \leq 1$ and $1-a^2-b^2-c^2+2abc \geq 0$, and $\lambda_1\lambda_2\lambda_3\lambda_4 \geq 0$.

Our examples are reminiscent of the examples considered in the setting of contextuality [30]. Note that our constraints are related, but nevertheless of a different flavor since we only consider such constraints for measurements which all commute. It may be interesting to consider such a moment matrix in order to determine how "non-contextual" quantum theory is. In section 4.1 and 3 we will develop classes of states which are restricted by imposing specific relationships among various moments. In particular, it will be of crucial importance whether we merely impose such constraints for measurements acting on different systems, or include such constraints for all commuting measurements.

3 p-nonlocal theories and their properties

We now define a series of so-called p-nonlocal "theories", each one more constrained than the previous. Our definition is thereby motivated by the uncertainty relations of [42] stated above. We later relate our definitions to Barrett's GNST [8] and what are commonly known as non-local

boxes. Our aim by constructing this series of simple theories is thereby merely to gain a more intuitive understanding of superstrong non-local correlations due to non-local boxes.

3.1 A theory without consistency constraints

We start with the simplest of all p-theories, which forms the basis of all subsequent definitions. In essence, we will simply allow states violating the uncertainty relation in 1 without worrying about anything else. In the spirit of Barrett [8] we start by defining the states which are allowed in our theory, and then allow all linear transformations preserving the set of allowed states. For simplicity, we will only consider the case of $d = 2^n$.

Definition 3.1. *A d-dimensional* p-bin state *is a $d \times d$ complex Hermitian matrix*

$$\rho = \frac{1}{d}\left(\mathbb{I} + \sum_{a,b} s_{ab} S_{ab}\right)$$

satisfying

1. *for all a, b, $-1 \leq s_{ab} \leq 1$.*
2. *for any set of mutually anti-commuting strings of Paulis $A_1, \ldots, A_m \in \mathbb{C}^{d\times d}$*

$$\sum_j |\operatorname{Tr}(A_j \rho)|^p \leq 1.$$

It remains to be specified what operations and measurements we are allowed to perform on p-bin states. We define

Definition 3.2. *A d-dimensional* p-bin theory *consists of*

1. *states $\rho \in \mathcal{S}_p^d$ where $\mathcal{S}_p^d$ is the set of d-dimensional p-bin states,*
2. *linear operations $T : \mathcal{S}_p^d \to \mathcal{S}_p^d$,*
3. *measurements described by observables $S_{ab} = S_{ab}^0 - S_{ab}^1$ where S_{ab}^0 and S_{ab}^1 are projectors onto the positive and negative eigenspace of S_{ab} respectively. As in the quantum case we let*

$$p_0 = \operatorname{Tr}(\rho S_{ab}^0) \text{ and } p_1 = \operatorname{Tr}(\rho S_{ab}^1).$$

Starting from a state, we may apply any set of operations T followed by a single measurement.

Note that by virtue of Eq. (1) any quantum state is a p-bin state. Note that the converse however does not hold, since the conditions given above do not imply that a p-bin state ρ is positive semi-definite. It seems very restrictive to limit ourselves to a single measurement at the end. The reason for this is that for some p, there exist p-bin states to start with, valid operations and measurements, followed by another operation that give us a states that are no longer a p-bin states [12]. We return to this question, when we consider the set of allowed operations below.

Note that the above definition is well-defined. First, we want that for any measurement S_{ab}, $\{p_0, p_1\}$ forms a valid probability distribution. A small calculation gives us that any p-nonlocal state ρ we have

$$p_v = \mathrm{Tr}(\rho S_{ab}^v) = \frac{1}{2}\left(1 + (-1)^v s_{ab}\right),$$

and thus $0 \leq p_b \leq 1$ and $p_0 + p_1 = 1$. Second, we want the non-signaling conditions to hold. When measuring $S_{ab} \otimes S_{a'b'}$ on a bipartite state

$$\rho_{AB} = \frac{1}{d}\left(\mathbb{I} + \sum_{\ell,m,\ell',m'} S_{\ell,m} \otimes S_{\ell',m'}\right)$$

we have that the probability to obtain outcome u for the measurement on the first system is given by

$$\Pr[u|ab, a'b'] = \sum_{v\in 0,1} \mathrm{Tr}\left(\rho_{AB}(S_{ab}^u \otimes S_{a'b'}^v)\right) = \frac{1}{2}(\mathbb{I} + (-1)^u s_{a,b,0,0}),$$

and hence $\Pr[u|ab, a'b'] = \Pr[u|ab, a''b'']$ for all a', b', a'', b'' as desired. A similar argument can be made to show that the more general independence condition is satisfied.

3.1.1 Basic Properties

We now state some basic properties of this theory, which will also hold for a more restricted p-nonlocal theory as outlined below.

Claim 3.3. *If ρ is a p-bin state, then ρ is also a q-bin state for $p, q \in \mathbb{Z}$ with $q \geq p$.*

Proof. This follows immediately from the fact that for any $r \in [0, 1]$ we have $r^q \leq r^p$. □

Below, we will apply circuits consisting of the Clifford gates $\{CNOT, X, Z, Y, H\}$ and $\mathbb{I}$. It is easy to see that such unitary operations are allowed transformations taking p-bin states to p-bin states.

Claim 3.4. *Let $\rho \in \mathcal{S}_p^d$. Then for any circuit U consisting solely of the gates $\{CNOT, X, Z, Y, H, \mathbb{I}\}$ we have $U\rho U^\dagger \in \mathcal{S}_p^d$.*

Proof. Note that U is composed of single unitaries $U_j = \mathbb{I}^{j-1} \otimes V \otimes \mathbb{I}^{n-j}$ with $V \in \{X, Z, Y, H\}$ and unitaries $U_j' = \mathbb{I}^{j-1} \otimes \mathrm{CNOT} \otimes \mathbb{I}^{n-j-1}$. First, it is straightforward to verify that for any $a, b \in \{0, 1\}^n$, there exist $a', b' \in \{0, 1\}^n$ such that $U_j S_{ab} U_j^\dagger = S_{a'b'}$, and similarly for U_j'. Second, applying a unitary to any set of anti-commuting operators again gives us anti-commuting operators. Hence, since we have $\sum_j |\mathrm{Tr}(A_j\rho)|^p \leq 1$ for *any* set of anti-commuting strings of Paulis, the resulting state will also have this property. □

It will also be useful to know that

Claim 3.5. *Let $\rho_1, \ldots, \rho_n \in \mathcal{S}_p^2$. Then $\bigotimes_{i=1}^n \rho_i \in \mathcal{S}_p^{2^n}$.*

Proof. We proceed by induction. By assumption, $\rho_1 \in \mathcal{S}_p^2$. We will show that for any states $\rho \in \mathcal{S}_p^{2^n}, \sigma \in \mathcal{S}_p^2$, the state $\rho \otimes \sigma \in \mathcal{S}_p^{2^{n+1}}$.

We need to prove that for any set of mutually anti-commuting Pauli's $A_j \in \mathbb{C}^{2^{n+1} \times 2^{n+1}}$ $\sum_j |\operatorname{Tr}(A_j \rho \otimes \sigma)|^p \leq 1$. Each A_j can always be written in terms of a Pauli, B_j acting on ρ, plus a Pauli $\{\mathbb{I}, X, Y, Z\}$ on σ. We separate the A_j into groups according to which Pauli is appended to B_j. Then we can rewrite this as

$$\begin{aligned}
&\sum_{j_{\mathbb{I}}} |\operatorname{Tr}((B_{j_{\mathbb{I}}} \otimes \mathbb{I})(\rho \otimes \sigma))|^p + \sum_{j_X} |\operatorname{Tr}((B_{j_X} \otimes X)(\rho \otimes \sigma))|^p \\
+ &\sum_{j_Y} |\operatorname{Tr}((B_{j_Y} \otimes Y)(\rho \otimes \sigma))|^p + \sum_{j_Z} |\operatorname{Tr}((B_{j_Z} \otimes Z)(\rho \otimes \sigma))|^p \\
&= \sum_{j_{\mathbb{I}}} |\operatorname{Tr}(B_{j_{\mathbb{I}}} \rho)|^p + \sum_{j_X} |\operatorname{Tr}(B_{j_X} \rho)|^p |\operatorname{Tr}(X\sigma)|^p \\
+ &\sum_{j_Y} |\operatorname{Tr}(B_{j_Y} \rho)|^p |\operatorname{Tr}(Y\sigma)|^p + \sum_{j_Z} |\operatorname{Tr}(B_{j_Z} \rho)|^p |\operatorname{Tr}(Z\sigma)|^p \leq 1
\end{aligned}$$

Since all the A_j mutually anti-commute, then for different j, j', $\{B_j \otimes X, B_{j'} \otimes X\} = 0$ implies $\{B_j, B_{j'}\} = 0$, while $\{B_j \otimes X, B_{j'} \otimes Y\} = 0$ implies $[B_j, B_{j'}] = 0$. Then because $\rho \in \mathcal{S}_p^{2^n}$ and $\{B_{j_X}, B_{j'_X}\} = 0$, and, for similar reasons $\{B_{j_X}, B_{j_{\mathbb{I}}}\} = \{B_{j'_{\mathbb{I}}}, B_{j_{\mathbb{I}}}\} = 0$, we know

$$\sum_{j_{\mathbb{I}}} |\operatorname{Tr}(B_{j_{\mathbb{I}}} \rho)|^p + \sum_{j_X} |\operatorname{Tr}(B_{j_X})|^p \leq 1$$

Now we will shorten our notation by writing

$$\begin{aligned}
a_X &= |\operatorname{Tr}(X\sigma)|^p & b_X &= \textstyle\sum_{j_X} |\operatorname{Tr}(B_{j_X} \rho)|^p \\
a_Y &= |\operatorname{Tr}(Y\sigma)|^p & b_Y &= \textstyle\sum_{j_Y} |\operatorname{Tr}(B_{j_Y} \rho)|^p \\
a_Z &= |\operatorname{Tr}(Z\sigma)|^p & b_Z &= \textstyle\sum_{j_Z} |\operatorname{Tr}(B_{j_Z} \rho)|^p \\
& & b_{\mathbb{I}} &= \textstyle\sum_{j_{\mathbb{I}}} |\operatorname{Tr}(B_{j_{\mathbb{I}}} \rho)|^p
\end{aligned}$$

This allows us to write inequalities implied by the uncertainty relation like:

$$\begin{aligned}
a_X + a_Y + a_Z &\leq 1 \\
b_X + b_{\mathbb{I}} &\leq 1 \\
b_Y + b_{\mathbb{I}} &\leq 1 \\
b_Z + b_{\mathbb{I}} &\leq 1
\end{aligned}$$

We can also see that $a_X, a_Y, a_Z, b_X, b_Y, b_Z, b_{\mathbb{I}} \geq 0$. The task at hand is to show that these inequalities imply the one required of a state in $\mathcal{S}_p^{2^{n+1}}$, which we can now rewrite as

$$a_X b_X + a_Y b_Y + a_Z b_Z + b_{\mathbb{I}} \leq 1.$$

We do this by writing down a sum of products of non-negative quantities like $1 - a_X - a_Y - a_Z$ and noting that the result is non-negative.

$$a_X(1 - b_X - b_{\mathbb{I}}) + a_Y(1 - b_Y - b_{\mathbb{I}}) + a_Z(1 - b_Z - b_{\mathbb{I}}) + (1 - b_{\mathbb{I}})(1 - a_X - a_Y - a_Z) \geq 0$$

That equation can be rewritten as $1 - (a_X b_X + a_Y b_Y + a_Z b_Z + b_{\mathbb{I}}) \geq 0$, which is what we set out to show. Therefore, $\rho \otimes \sigma$ is a valid state, and, by induction, so is $\bigotimes_{i=1}^n \rho_i \in \mathcal{S}_p^{2^n}$ for any n. □

3.2 An analogue to box-world

Note that in the above definition we have not placed any constraints at all on the expectation values of commuting measurements. This was not necessary, as we had allowed a single measurement only, where by the above definition $\mathbb{I} \otimes X$ formed such a single measurement. Now consider a two-qubit system, i.e., $d = 4$. Suppose that we have for a particular ρ that

$$\mathrm{Tr}\left((X \otimes \mathbb{I})\rho\right) = \mathrm{Tr}\left((\mathbb{I} \otimes X)\rho\right) = \mathrm{Tr}\left((X \otimes X)\rho\right) = -1.$$

Note that ρ can be a perfectly valid state with respect to the definition given above, but yet we would not consider this to be consistent behavior, if we were allowed to perform subsequent measurements. We now introduce additional constraints that eliminate this inconsistency. It should be clear from section 2.3 that that to achieve full consistency we would have to introduce certain constraints for commuting observables in general. Yet, we will first restrict ourselves to observables on different systems in analogy to "box-world". We will show in section 4.1 that Barrett's GNST and non-local boxes essentially correspond to this definition. We will also see in section 6 and 7.1 that these additional constraints play a crucial role in the power of our model with respect to information processing tasks.

Definition 3.6. *A* p-box state *is a* p*-bin state* ρ*, where in addition we require that for any set* $C \in \mathcal{E}_L$ *of measurements acting on different systems and* $s(C)$ *as defined in Eq. (5) we have that the corresponding moment matrix* K_s *defined in section 2.4 satisfies*

$$K_s \geq 0.$$

Note that claims 3.3 and 3.5 holds analogously for p-box states. It is important to note though that claim 3.4 does not hold in this case, since for example the CNOT operation can lead to states violating the definition.

3.3 A theory with consistency constraints

Finally, we will impose all constraints required from our consistency considerations of section 2.3.

Definition 3.7. *A* p-nonlocal state *is a* p*-box state* ρ*, where in addition we require that for any set of commuting measurements* $C \in \mathcal{E}_C$ *and* $s(C)$ *as defined in Eq. (5) we have that the corresponding moment matrix* K_s *as defined in section 2.4 satisfies*

$$K_s \geq 0.$$

Again claims 3.3 and 3.5 hold analogous to the above. When we include all consistency considerations, it is also easy to see that claim 3.4 holds for p-nonlocal states, since for any allowed unitary U we already have by the above that ρ satisfies the constraints given by the set $C' = \{U^\dagger M_1 U, \ldots, U^\dagger M_m U\}$ and hence $U\rho U^\dagger$ remains a valid p-non-local state.

4 Generalized non-local theories

To create a closer analogy between our "theories" derived from relaxed uncertainty relations and non-local boxes, we now consider a related class of theories called *generalized no-signaling theories*

(GNST) [8], for which we will consider similar relaxations. As already sketched in the introduction, states in a GNST are defined operationally. Consider a laboratory setup where we have a device which prepares a specific state. We then use a measuring device which has a choice of settings allowing us to measure different properties of the system. The measuring device gives us a reading specifying the outcome of the measurement. A particular state in GNST is described completely by means of the probabilities of obtaining each outcome when performing a fixed set of *fiducial* measurements. For example, for a set of fiducial measurements $\mathcal{O} = \{X, Z, Y\}$ with outcomes $\mathcal{A} = \{\pm 1\}$, the probabilities $p(A|C)$ for all $A \in \mathcal{A}$ and $C \in \mathcal{O}$ form a description of the state. Hence, we will simply use p to refer to a state given by said conditional probabilities. The idea behind considering fiducial measurements stems from the idea that there exists a set of measurement choices that suffice to fully describe the system. In classical mechanics, for instance, we can always in principle make a single measurement which outputs all the information necessary to describe a state. For a qubit, on the other hand, we would need results from at least three different incompatible measurement settings, e.g., spin in three orthogonal directions. We refer to [8] for a definition of GNST and its allowed operations. For us it will only be important to note that similar to the setting of non-local boxes, we can make only one measurement on each system, and there is no real notion of post-measurement states defined.

In the following, we will be interested in the special case of multi-partite systems where on each system we can perform one of three fiducial measurements with outcomes ± 1. Using our notation from section 2.2 we write the set of realizable experiments for GNST as

$$\mathcal{E}_G = \{\forall k \in \{1,2,3\}^n : \{W_{1,k_1}, \ldots, W_{n,k_n}\}\},$$

with W_{i,k_i} denoting a choice of the k_ith measurement on the ith system. Later we will connect these measurement choices with Pauli measurements via the relation $W_{i,1} = X_i, W_{i,2} = Z_i, W_{i,3} = X_i Z_i$. A key point of this definition will be that the partitioning of measurements into n systems will be fixed. We also demand that probability distributions should satisfy an independence principle. As we pointed out, when restricted to partitions over disjoint parties, this just reduces to the no-signaling principle. That is, the choice of measurement on one subset of particles can not be used to send a signal to a disjoint subset.

In analogy to the quantum setting [8], we let one gbit refer to a single system on which we can perform our set of fiducial measurements given above. Our definition of a gbit thereby slightly differs from the definition given in [8], which only allows two fiducial measurements X and Z on a single gbit. Yet, in order to compare the hierarchy of GNST-like theories we will construct below to the p-box states from above we adopt this slightly more general definition in analogy to a single qubit in the quantum case. Note that for the set of measurements $C \in \mathcal{E}_G$ specified above, an n-gbit state, specified by $p : \mathcal{A}^{\times n} \times C \to [0,1]$, is in GNST if p satisfies constraints (1), (2), and (3') in section 2.2.

Example 4.1. *Consider the following state of one particle in GNST (or one gbit):*

$$\begin{array}{ccccc} p(A=+1|M=X) & = & s_x & = & 1 - p(A=-1|M=X) \\ p(A=+1|M=Z) & = & s_y & = & 1 - p(A=-1|M=Z) \\ p(A=+1|M=XZ) & = & s_z & = & 1 - p(A=-1|M=XZ) \end{array}$$

This state is normalized, and positivity requires $s_x, s_y, s_z \in [0,1]$. The state would be equivalent to the state of an arbitrary qubit if and only if $s_x^2 + s_y^2 + s_z^2 \leq 1$, that is, if we are constrained to the Bloch sphere.

For multi-partite states the difference between constraints on qubits and gbits becomes more complicated. We now turn to describing a hierarchy of constraints on GNST theories which will be analogous to uncertainty conditions in p-nonlocal theories and quantum mechanics.

4.1 p-GNST

Even though states in GNST are defined without any particular structure to their measurements embedded, we will now impose a physically motivated structure. In particular, we will simply *imagine* in analogy to the quantum setting that measurements X, Z and Y obey the same anti-commutation relations as the Pauli matrices $\{X, Z\} = \{Z, Y\} = \{X, Y\} = 0$. In our definition below, we will for simplicity write $\{\cdot, \cdot\}$ to indicate that we imagine such an anti-commutation constraint to hold exactly when the string of Paulis $\prod_i W_{i,k_i}$ associated with each C would anti-commute.

First of all, this will allow us to artificially impose an uncertainty relation just like Eq. (1).

Definition 4.2. *A state is in p-GNST if it is in GNST and for any set of measurements $S = \{C \in \mathcal{E}_G\}$ satisfying that for all $C, C' \in S$, $\{C, C'\} = 0$ we have*

$$\sum_{C \in S} |m(C)|^p \leq 1. \tag{6}$$

Note that for $p \to \infty$ this condition no longer restricts the states, because we get $\max_{C \in \mathcal{S}} |m(C)| \leq 1$, which is true for the original GNST, and non-local boxes. If we would actually add such commutation and anti-commutation constraints we could now again distinguish between adding the consistency constraints of section 2.3 only for measurements acting on different systems, or for all commuting measurements in analogy to the p-box and p-nonlocal theories. In analogy to GNST, where commutation relations were only defined for measurements acting on different systems however, we will stick to this setting, even when considering $p < \infty$. A p-GNST state is thus essentially analogous to a p-box state, except we are allowed to make simultaneous measurements of locally disjoint systems.

5 Superstrong non-locality

Before we show that relaxing the uncertainty equation of Eq. (1) leads to superstrong non-local correlations, let's take a look at what effect this uncertainty relation actually has on quantum strategies for the CHSH inequality. For this purpose, we will rewrite Tsirelson's bound for the CHSH inequality in its more common form as

$$|\langle A_0 \otimes B_0 \rangle + \langle A_0 \otimes B_1 \rangle + \langle A_1 \otimes B_0 \rangle - \langle A_1 \otimes B_1 \rangle| \leq 2\sqrt{2},$$

where we use A_0, A_1 and B_0, B_1 to denote Alice's and Bob's observables respectively where $A_0^2 = A_1^2 = B_0^2 = B_1^2 = \mathbb{I}$. We will use the fact that in order to achieve the maximum possible quantum violation we must have $\{A_0, A_1\} = 0$ and $\{B_0, B_1\} = 0$ [14, 38, 34]. For $M_1 = A_0 \otimes B_0$, $M_2 = A_0 \otimes B_1$, $M_3 = A_1 \otimes B_0$ and $M_4 = A_1 \otimes B_1$ this means that we have $\{M_1, M_2\} = \{M_1, M_3\} = \{M_2, M_4\} = \{M_3, M_4\} = 0$. Using the uncertainty relation of Eq. (1) proving Tsirelson's bound is equivalent to solving the following optimization problem

$$\begin{array}{ll} \text{maximize} & \langle M_1\rangle + \langle M_2\rangle + \langle M_3\rangle - \langle M_4\rangle \\ \text{subject to} & \langle M_1\rangle^2 + \langle M_2\rangle^2 \leq 1 \\ & \langle M_1\rangle^2 + \langle M_3\rangle^2 \leq 1 \\ & \langle M_2\rangle^2 + \langle M_4\rangle^2 \leq 1 \\ & \langle M_3\rangle^2 + \langle M_4\rangle^2 \leq 1 \end{array}$$

By using Lagrange multipliers, it is easy to see that for the optimum solution we have $\langle M_1\rangle^2 = \langle M_4\rangle^2$ and $\langle M_2\rangle^2 = \langle M_3\rangle^2$. By considering all different possibilities, we obtain that with $x = \langle M_1\rangle = -\langle M_4\rangle$ and $y = \langle M_2\rangle = \langle M_3\rangle$ our optimization problem becomes

$$\begin{array}{ll} \text{maximize} & 2(x+y) \\ \text{subject to} & x^2 + y^2 \leq 1 \end{array}$$

Again using Lagrange multipliers, we now have that the maximum is attained at $x = y = 1/\sqrt{2}$ giving us Tsirelson's bound.

Tsirelson's bound can hence be understood as a consequence of the uncertainty relation of [42]. Thus, we intuitively expect that relaxing this relation affects the strength of non-local correlations. In a similar way, one can view monogamy of non-local correlations as a consequence of Eq. (1) [35].

5.1 CHSH inequality

5.1.1 In p-theories

To see what is possible in p-theories, we first construct the equivalent of a maximally entangled state. Let

$$\rho_p = \frac{1}{2}\left[\mathbb{I} + \left(\frac{1}{2}\right)^{\frac{1}{p}}(X+Y)\right].$$

Note that for $p \to \infty$ this gives us

$$\rho_\infty = \frac{1}{2}\left[\mathbb{I} + X + Y\right].$$

We now proceed analogously to the quantum case to construct

$$\eta_1 = \mathrm{CNOT}(\rho_p \otimes |0\rangle\langle 0|)\mathrm{CNOT}^\dagger,$$

which by claim 3.4 is a valid p-bin and p-nonlocal state. It can also be verified that η_1 forms a valid p-box state.

Claim 5.1. *Let $A_1 = X$, $A_2 = Y$, $B_1 = X$ and $B_2 = Y$ be Alice and Bob's observables respectively. Then*

$$\langle CHSH_p\rangle = \mathrm{Tr}(\eta_1(A_1\otimes B_1 + A_1\otimes B_2 + A_2\otimes B_1 - A_2\otimes B_2)) = 4\frac{1}{2^{1/p}},$$

for all p-theories.

Proof. This follows immediately by noting that

$$\eta_1 = \frac{1}{4}\left(\mathbb{I} + \frac{1}{2^{1/p}}(X\otimes X + X\otimes Y + Y\otimes X - Y\otimes Y) + Z\otimes Z\right).$$

□

We can also phrase this statement in terms of probabilities as stated in the introduction, by noting that the maximum probability that Alice and Bob win the CHSH game is given by

$$\frac{1}{2} + \frac{\langle CHSH_p \rangle}{8} = \frac{1}{2} + \frac{1}{2 \cdot 2^{1/p}}.$$

It is important to note that this violation can be obtained even when imposing the additional consistency constraints from section 2.3.

5.1.2 In p-GNST

We already saw in the introduction that GNST admits states analogous to a non-local box, allowing for a maximal violation of the CHSH inequality. We now show that similar states exist for p-GNST theories analogous to p-box states. We first phrase the CHSH inequality in terms of probabilities. In particular, consider the GNST state specified by $p((A_1, A_2)|\{M_1, M_2\}) = \frac{1}{4}(1 + (-1)^{\delta_{M_1,Z_1}\delta_{M_2,Z_2}} A_1 A_2 \lambda)$ for some λ to be chosen below. If each party measures X or Z on their state and outputs the result ± 1, the probability that Alice and Bob win the CHSH game is given by

$$\begin{aligned}&\frac{1}{4}(p(1,1|X_1,X_2) + p(-1,-1|X_1,X_2) + p(1,1|X_1,Z_2) + p(-1,-1|X_1,Z_2) \\ &+p(1,1|Z_1,X_2) + p(-1,-1|Z_1,X_2) + p(1,-1|Z_1,Z_2) + p(-1,1|Z_1,Z_2)) = \frac{1+\lambda}{2}\end{aligned}$$

In terms of the moments, $m(X_1, X_2) = m(X_1, Z_2) = m(Z_1, X_2) = -m(Z_1, Z_2) = \lambda$, and this becomes

$$\frac{1}{4}(2 + \frac{1}{2}(m(X_1,X_2) + m(X_1,Z_2) + m(Z_1,X_2) - m(Z_1,Z_2))) = \frac{1+\lambda}{2}$$

Now we can consider the maximum value of λ that is a valid state in p-GNST. The requirements listed in example 2.3 only restrict $|\lambda| \leq 1$. Eq. (6) requires $|m(X_1, X_2)|^p + |m(X_1, Z_2)|^p = |m(Z_1, X_2)|^p + |m(Z_1, Z_2)|^p = 2|\lambda|^p \leq 1 \rightarrow \lambda = (\frac{1}{2})^{\frac{1}{p}}$. Therefore in a p-GNST it is possible to win the CHSH game with probability $1/2 + 1/(2 \cdot 2^{1/p})$.

5.2 XOR games

We now investigate the case of general 2-player XOR-games for $p \rightarrow \infty$. In such a game we have an arbitrary (but finite) set of questions S and T from which Alice's and Bob's questions $s \in S$ and $t \in T$ are chosen according to a fixed probability distribution $\pi : S \times T \rightarrow [0, 1]$. Yet, the set of possible answers remain $A = B = \{0, 1\}$ for Alice and Bob respectively. The game furthermore specifies a predicate $V : A \times B \times S \times T \rightarrow \{0, 1\}$ that determines the winning answers for Alice and Bob. In an XOR game, this predicate depends only on the XOR $c = a \oplus b$ of Alice's answer a and Bob's answer b. We thus write $V(c|s, t) = 1$ if and only if answers $a \oplus b$ satisfying $a \oplus b = c$ are winning answers for questions s and t. We will also restrict ourselves to unique games, which have the property that for any s, t, b, there exists exactly one winning answer a for Alice (and similarly for Bob).

First of all, note that in the quantum case we may write the probability that Alice and Bob return answers a and b with $a \oplus b = c$ as

$$p(c|s,t) = \frac{1}{2}(1 + (-1)^c \langle \Psi | A_s \otimes B_t | \Psi \rangle),$$

where we again use A_s and B_t to denote Alice's and Bob's observable corresponding to questions s and t respectively and $|\Psi\rangle$ denotes the maximally entangled state. Note that we again have $(A_s)^2 = (B_t)^2 = \mathbb{I}$ from the fact that both measurements have only two outcomes. The probability that Alice and Bob win the game can then be written as

$$\sum_{s,t} \pi(s,t) \sum_{c} V(c|s,t) p(c|s,t).$$

Let $v_{st} = \langle \Psi | A_s \otimes B_t | \Psi \rangle$. First of all note that for $p \to \infty$

$$\frac{1}{d} \left(\mathbb{I} + \sum_{st} v_{st} \Gamma_s \otimes \Gamma_t \right) \tag{7}$$

with $d = 2^{\max |S|, |T|}$ and Γ_s, Γ_t anti-commuting observables as defined in section 2 is a valid state for any $|v_{st}| \leq 1$. Hence, we can immediately see that

Corollary 5.2. *In any ∞-theory, there exists a strategy for Alice and Bob to win a unique XOR game with certainty.*

Proof. Consider the state given in Eq. (7) with $v_{st} = \pm 1$ such that $p(c|s,t) = 1$ whenever $V(c|s,t) = 1$. Let Alice and Bob's measurements be given by Γ_s and Γ_t for questions s and t respectively, which are valid measurements for all p-theories with Γ_s, Γ_t constructed as in section 2. □

We leave it as an open question to examine the case of $p < \infty$ for XOR games, since our aim was merely to show that superstrong correlations can exist, if we allow for relaxed uncertainty relations. We can see that letting $v_{st} = \pm 1/(\max |S|, |T|)^{1/p}$ makes Eq. (7) a valid state for any choice of p, but this may not generally be the optimal choice. The case of GNST is similar, and it has been shown that any non-local correlations can (approximately) be simulated by such boxes [23]. Optimal bounds for p-GNST with $p < \infty$ can be obtained using techniques analogous to [22].

6 Superstrong random access encodings

The existence of superstrong non-local correlations is by no means the only difference we can observe when moving from quantum theory to p-GNST or p-nonlocal theories. In particular, we now show that we can obtain so-called random access encodings which, depending on the theory, can be exponentially better than those realized by quantum mechanics. We then investigate how uncertainty relations and the restrictions imposed by simultaneous measurements affect this encoding. The existence of such random access encodings will play a crucial role when considering the power of p-GNST theories for communication complexity in section 7.1. In section 7.2 we also use this random access code to prove a lower bound on the sample complexity of learning states in GNST.

6.1 In p-GNST

Intuitively, a random access code [2, 3] allows us to encode N bits into a physical system of size n such that we can decode any one bit of the original string with probability at least q. More formally,

Definition 6.1. *A $[N, n, q]$-random access code (RAC) is an encoding of a string $x \in \{0,1\}^N$ into an n-gbit state p_x, such that there exist measurements $C \in \mathcal{E}_G$ with outcomes $A \in \mathcal{A}^{\times n}$, and a decoding algorithm $D : \mathcal{A}^{\times n} \to \{0,1\}$ satisfying*

$$Pr(D(A) = x_k) = \sum_{A \in \mathcal{A}^{\times n}} \delta_{D(A),x_k} p_x(A|C) \geq q,$$

where $p_x(A|C)$ is the probability of obtaining outcome A when performing the measurement C.

It has been shown [2, 3] that in the quantum case, we must have $n \geq (1 - h(q))N$, where h denotes the binary entropy function. There also exist classical encodings for which $n = (1 - h(q))N + O(\log N)$ [2]. Hence, quantum states offer at most a modest advantage over classical mechanics and, for $q = 1$, no advantage at all. We now proceed to the surprising result that general no-signaling states lead to extremely powerful random access codes.

Claim 6.2. *In GNST, there exists a $[3^n, n, 1]$-random access code.*

Proof. An n gbit state in GNST is completely characterized by the probabilities of outcomes for a fixed set of measurements. Recall that a single gbit is a two-level system on which we allow three possible measurements with two possible outcomes each. Also recall that each $C \in \mathcal{E}_G$ can be represented as $\mathcal{E}_G = \{\forall k \in \{1,2,3\}^n : \{W_{1,k_1}, \ldots, W_{n,k_n}\}\}$, with $W_{i,1} = X_i, W_{i,2} = Z_i, W_{i,3} = X_i Z_i$. Note that each measurement C is associated with one of $N = 3^n$ vectors $k = (k_1, \ldots, k_n)$. Let $f : C \to \{1, \ldots, N\}$ be a one-to-one function. For each of the $N = 3^n$ bits we wish to encode, we must specify one measurement C that we can use to extract the jth-bit. Let that measurement be denoted by $f^{-1}(j)$.

We are now ready to define our encoding of the string $x \in \{0,1,2\}^N$ into an n-gbit GNST state p_x via the probabilities

$$p_x(A|C) := \frac{1}{2^n}(1 + A^*(-1)^{x_{f(C)}}),$$

where we use the previously defined notation $A^* = \prod_{i=1}^{|C|} A_i$. It is straightforward to verify that the state is normalized, positive, and satisfies the no-signaling condition.

We now show that any bit of the original string can be decoded perfectly. If we choose to retrieve bit j, we measure $C = f^{-1}(j)$. That means that we get result A with probability $\frac{1}{2^n}(1+A^*(-1)^{x_j}) = \frac{1}{2^n} 2\delta_{A^*,(-1)^{x_j}}$. And we get the result $A^* = (-1)^{x_j}$ with probability:

$$\sum_{A^*=(-1)^{x_j}} p_x(A|C) = \sum_{A^*=(-1)^{x_j}} \frac{1}{2^n} 2\delta_{A^*,(-1)^{x_j}} = 1.$$

where the last equality follows from the fact that we sum over exactly half the 2^n possible outcomes $A_1, \ldots, A_n$. Hence the decoder $D(A) = \frac{1}{2}(1 - A^*)$ will return x_j with perfect probability. □

What happens if we impose the uncertainty relation in p-GNST? For convenience sake, note that we could rewrite the encoding above in terms of moments, where we let an encoding of a string x be determined by the moment representation of p_x as

$$m_x(C = f^{-1}(k)) := (-1)^{x_k}$$

with all other moments set to 0.

To construct an encoding for p-GNST, we consider

$$m_x(C = f^{-1}(k)) := (-1)^{x_k}\lambda.$$

What's the largest λ that satisfies the uncertainty relation? As we noted earlier the maximum number of anti-commuting Pauli operators is $2n+1$, so the most restrictive condition we could get from the uncertainty relation is $(2n+1)|\lambda|^p \leq 1$. We thus obtain

Claim 6.3. *In p-GNST, there exists a $[3^n, n, \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2n+1}\right)^{1/p}]$-random access code.*

Proof. Let $\lambda = (2n+1)^{1/p}$, and note that this satisfies the uncertainty relation. Our encoding is now

$$p_x(A|C) = \frac{1}{2^n}(1 + (-1)^{x_{f(C)}}\ \lambda\ A^*).$$

And our probability of getting the correct sign from our measurement goes down to

$$\Pr(D(A) = x_k) = \frac{1+|\lambda|}{2} = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2n+1}\right)^{1/p}$$

□

If $p < \infty$ we get an encoding that gets asymptotically worse for large n. This should be compared to the bound on the number of qubits for a *quantum* random access encoding of $N = 3^n$ bits into k qubits with recovery probability $q = 1/2 + 1/2(1/(2n+1))^{1/p}$. From the bound of [2, 3], we have that the encoding uses exponentially fewer physical bits than what can be obtained in the quantum setting and hence even p-GNST has a powerful coding advantage over quantum mechanics. Note that we are always free to split the N bits into smaller pieces first, and encode each piece independently to keep the recovery probability q constant. This is analogous to the quantum setting where we can encode each 3 bits into one qubit to obtain a random access code with $n = N/3$. Alternatively, we can form a simple repetition code, where we have k copies of the random access codes constructed above. We then have

Claim 6.4. *In p-GNST, there exists a $[3^n, (2n+1)^{3/p}n, 1-\varepsilon]$-random access code with $\varepsilon = 2\exp(-(2n+1)^{1/p}/2)$.*

Proof. We take k copies of the RAC defined in Claim 6.3, and decode by taking the majority of the individual encodings. Let $Y_j = 1$ if the decoding was successful for the j-th copy, and $Y_j = 0$ otherwise. From the Hoeffding inequality we immediately obtain that for $Y = \sum_{j=1}^{k} Y_j$ and q as defined above

$$\Pr\left[|Y - qk| \geq t\ k\right] \leq 2e^{-2t^2 k},$$

If we set $t = q - 1/2 = 1/2(1/(2n+1))^{1/p}$, that gives us $\Pr\left[Y \leq k/2\right] \leq 2e^{-\frac{1}{2}\left(\frac{1}{2n+1}\right)^{2/p}k}$. Now if we set $k = (2n+1)^{3/p}$, we have used a total of $(2n+1)^{3/p}n$ gbits and will succeed with probability $1 - 2e^{-(2n+1)^{1/p}/2}$ as promised. □

Whereas $(2n+1)^{3/p}n$ is still quite large, note that it is nevertheless only polynomial in n. The length of the RAC is hence still poly-logarithmic in our original input size, where we achieve (near) perfect recovery for large n. Finally, we will need to use one more related result.

Claim 6.5. *In p-GNST, for $\gamma \in (0, 1/2)$ and $\hat{n} \geq 2^{2/p} \ln(4/(1/2-\gamma)^2)$, there exists a $[3^{n(\hat{n},p,\gamma)}, \hat{n}, \frac{1}{2}+\gamma]$-random access code with $n(\hat{n}, p, \gamma) = \lfloor \left(\frac{\hat{n}\ 2^{-2/p}}{\ln(4/(1/2-\gamma)^2)} \right)^{\frac{1}{2/p+1}} \rfloor$.*

Proof. Again we take k copies of the RAC defined in Claim 6.3, and decode by taking the majority of the individual encodings. The probability to decode correctly in that case was $1-2e^{-\frac{1}{2}(\frac{1}{2n+1})^{2/p}k}$. Now we want to adjust k and n to get a code with a fixed success rate and that uses no more than $\hat{n}$ gbits. We need that (i) $kn \leq \hat{n}$, that is, our encoding uses at most $\hat{n}$ physical bits and (ii) $1 - 2e^{-\frac{1}{2}(\frac{1}{2n+1})^{2/p}k} \geq 1/2 + \gamma$, which forces our probability of success to be at least $1/2 + \gamma$. We can satisfy (ii) if we set $k = \ln(4/(1/2 - \gamma)^2)(2n + 1)^{2/p}$, then (i) tells us that $kn = \ln(4/(1/2 - \gamma)^2)(2n + 1)^{2/p}n$, from which we have $\ln(4/(1/2 - \gamma)^2)2^{2/p}n^{2/p+1} \leq kn \leq \hat{n}$ and thus

$$n \leq \left(\frac{\hat{n}\ 2^{-2/p}}{\ln(4/(1/2 - \gamma)^2)} \right)^{\frac{1}{2/p+1}}.$$

Since the smallest system we can encode into is $n = 1$, this tells us that $\hat{n}$ must be at least $2^{2/p} \ln(4/(1/2 - \gamma)^2)$. □

Note that although this may not be the best encoding, it suffices to give us the asymptotic behavior for $\hat{n}$.

6.2 In p-nonlocal theories

It is instructive to consider such superstrong encodings in the language of p-nonlocal theories to see how such superstrong encodings would look like in terms of Pauli matrices. This will also allow us to compare the consequences of restrictions due to the consistencies of moments from section 2.3 to random access encodings. For the least restrictive p-theory, the p-bin theory, we can construct the following very simple encoding.

Claim 6.6. *In p-bin theories, there exists a $[2^{2n} - 1, n, \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2n+1}\right)^{1/p}]$-random access code.*

Proof. Consider the encoding of a string $x \in \{0, 1\}^N$ with $N = 2^{2n} - 1$ into an n p-bit state given by

$$\rho_x := \frac{1}{d}\left(\mathbb{I} + \frac{1}{(2n + 1)^{1/p}} \sum_{k=1}^{2^{2n}-1} (-1)^x_k S_k \right),$$

where $S_k = S_{ab}$ is a string of Pauli matrices, where we simply relabeled the indices ab. To decode the kth-bit, we measure S_k. A straightforward calculation shows that the probability to obtain outcome x_k is given by

$$\Pr[x_k] = \frac{1}{2} \operatorname{Tr}\left[(\mathbb{I} + S_k)\, \rho_x \right] = \frac{1}{2} + \frac{1}{2(2n + 1)^{1/p}},$$

as promised. Clearly, the uncertainty relation is satisfied. □

Similarly, we obtain the following encoding for p-box theories, which is in one-to-one correspondence with the encodings in p-GNST above.

Claim 6.7. *In p-box theories, there exists a* $[3^n, n, \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2n+1}\right)^{1/p}]$*-random access code.*

Proof. Our encoding is analogous to the one above, but we restrict ourselves to including only such strings of Pauli matrices formed by taking tensor products of $\{X, Y, Z\}$, excluding the identity. □

Clearly, we can again obtain an encoding that is poly-logarithmic in the length of the original input analogous to Claim 6.4 that has perfect recovery for large n.

6.3 The effect of consistency

When viewing such encodings in terms of density matrices, it becomes clear why such encodings do not exist in a quantum setting: all such encodings are in gross violation of the consistency conditions of section 2.3. Even when we restrict ourselves to $p = 2$, we can obtain such encodings whereas in the quantum case we cannot. It is interesting to note that for $p = 2$, the violation we can obtain for e.g. the CHSH game is exactly the same as in the quantum setting. Thus it is perfectly possible to have such superstrong encodings, while simultaneously being restricted to Tsirelson's bound in the CHSH game for a 2 qubit state. This clearly shows how limited our p-bin, p-nonlocal, but also p-GNST theories really are. Since GNST is equivalent to a theory based on non-local boxes, this also shows that considering such boxes is somewhat limiting, and possibly ignores some aspect present in quantum theory that are of importance for information processing.

7 Implications for information processing

We now turn to a number of interesting implications of p-GNST and p-theories to information processing. In particular, we will see that both allow us to save significantly on the amount of data we need to transmit to solve certain communication problems. In fact, we will see that there exists a task for which there exists an *exponential* gap between the amount of communication required when compared with quantum theory. Other information tasks on the other hand become more difficult. We will see that when trying to learn states approximately we need to perform exponentially more measurements in the case of GNST.

7.1 Communication complexity

Imagine two (or more) parties, Alice and Bob, who each have an input $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ respectively, unknown to the other party. Their goal is to compute a fixed function $f : \{0,1\}^{2n} \to \{0,1\}^m$ by communicating over a channel. The central question of communication complexity is how many bits they need to transmit in order to compute f. Typically, we thereby only require one party (Bob) to learn the result $f(x, y)$. To help them reduce the amount of communication, Alice and Bob may possess additional resources such as shared randomness, entanglement, non-local boxes or communicate over a quantum channel, and may impose different measures of success. For example, they could be interested in computing f only with a certain probability instead of computing it exactly. It is well-known that if Alice and Bob can share non-local boxes, they can compute any Boolean function $f : \{0,1\}^{2n} \to \{0,1\}$ perfectly by communicating only a single bit [39], which is even true when the non-local boxes have slight imperfections [11]. Here, we consider the case where Alice and Bob have *no* a-priori resources, however, we they are able to exchange p-GNST or p-nonlocal states over a suitable channel.

7.1.1 One-way communication

We first of all make a very modest statement and show that in *any* one-way communication protocol, where Alice sends a single message to Bob, we are able to save a constant number of bits, when computing a Boolean function f. These savings are an immediate consequence of the existence of superstrong random access codes that we discussed in section 6. To communicate with Bob, Alice constructs the string

$$m = f(x,0), \ldots, f(x,2^n-1)$$

and encodes $m \in \{0,1\}^{2^n}$ into a random access code ρ_m. To retrieve the correct answer, Bob simply retrieves bit $x_y = f(x,y)$ from ρ_m. Evidently, this type of saving is particularly interesting in the case where Alice and Bob would need to communicate n bits to compute f, which is the case classically and quantumly if $f = IP$ is the inner product [20]. By Claims 6.2, 6.3, 6.7 and 6.6 we immediately obtain that

Claim 7.1. *Let $p \to \infty$. Then in to compute the inner product Alice needs to transmit at most k bits to Bob, where*

$$k = \begin{cases} (1/\log 3)n & \text{for } p\text{-GNST and } p\text{-nonlocal theories} \\ n/2 & \text{for } p\text{-bin theory} \end{cases}$$

7.1.2 Private information retrieval

More striking though are the possibilities of p-GNST or p-theories for the task of private information retrieval: Here, one (or more) database servers each hold a copy of the database string $x \in \{0,1\}^n$. A database user should be able to retrieve any bit x_i of his choosing, while the servers should not learn the desired index i. A protocol that satisfies these parameters is the trivial one, where the server simply sends the entire string x to the user. The question is thus, whether it is possible to perform this task by communicating less than n bits. If only a single server is used, it is known that the trivial protocol is optimal and we need to communicate $\Theta(n)$ bits, even if we are allowed quantum communication [25]. It is clear that the superstrong encodings from above, allow us to beat this bound trivially, by asking the server to encode x into a superstrong random access code. Hence we have as an immediate consequence of Claims 6.2, 6.4, 6.6, and 6.7 we have

Claim 7.2. *In any p-GNST, p-bin, and p-box theory, there exists a single server private information retrieval scheme requiring $O(polylog(n))$ bits of communication for large n.*

7.2 Learnability

We consider a scenario in which there is an unknown state for which we are trying to learn an approximate description. In particular, imagine some arbitrary probability distribution over possible two-outcome measurements. We are given the expectation value for each measurement in a finite set picked according to this distribution. We then construct an approximate description of the state which agrees with all the expectation values we have observed so far. This description is considered to be good if it predicts the correct results for most future measurements drawn from the same distribution. The central question is how many measurement results we need to be able construct a good description.

The existence of strong random access codes has implications for state learning. Aaronson [1] used an upper bound on the number of bits that can be encoded into an n qubit RAC to upper

bound the number of measurements needed to learn an approximate description of an n qubit state. He took solace in the fact that, despite the exponential number of parameters describing a quantum state, a linear (in the number of qubits) number of measurements suffice to learn an approximate description of the state. If an exponential number of measurements were really required, we could never hope to do enough measurements to verify the identity of quantum states of a few hundred particles.

We show the converse for states in p-GNST. We use our constructions of random access codes to lower bound the number of measurements needed to learn an approximate description of the state. We find that an exponential number of measurements is required to find such a description and therefore one could never hope to do enough measurements to learn a description of a state with a modest number of particles, even approximately. This holds even for theories where $p = 2$ and the violation of the CHSH inequality is the same as for quantum mechanics. This demonstrates an unusually powerful theory which starkly contrasts with quantum mechanics and the p-nonlocal theory.

We begin with a section defining the relevant tools: a definition of the learning scenario, and a measure of state complexity known as the "fat shattering dimension." We then restate a known lower bound on the number of samples needed for learning in terms of the fat shattering dimension. In the next section, we derive lower bounds on learnability for p-GNST theories. First, we use our random access codes to lower bound the fat shattering dimension for p-GNST states. Then we can use this result to lower bound the number of samples needed to learn p-GNST states.

7.3 Tools

We begin by introducing some terminology from statistical learning theory. Let the set $\mathcal{S}$ denote the sample space, which will correspond to the space of possible measurements in our case. A probabilistic concept over $\mathcal{S}$ is just a function $F : \mathcal{S} \to [0, 1]$, and is equivalent to a state which maps measurement choices to expectation values. A set of such concepts is referred to as the concept class $\mathcal{C}$ over $\mathcal{S}$ and corresponds to the set of all states. We consider the learning situation in which you are given the value of the target concept (state) over some samples drawn independently according to an arbitrary distribution. The goal is to output a hypothesis concept that will give values close to the target concept for most samples drawn from the same distribution. A sample size that is large enough to allow this to be accomplished with high probability is said to be sufficient. To restate the connection, in GNSTs we will say that a state corresponds to a concept, and a measurement on the state to a sample. We will make these notions precise before we demonstrate the connection between RACs and fat-shattering dimension in 7.5.

We adopt our definition of probabilistic concept learning from Anthony and Bartlett[4].

Definition 7.3 (Anthony and Bartlett [4])**.** *Let $\mathcal{S}$ be a sample space, let $\mathcal{C}$ be a probabilistic concept class over $\mathcal{S}$, and let $\mathcal{D}$ be a probability measure over $\mathcal{S}$. Fix an element $\rho \in \mathcal{C}$, as well as error parameters $\varepsilon, \eta, \gamma > 0$ with $\gamma > \eta$. Let $k_0(\eta, \gamma, \epsilon, \delta)$ be some function of the error parameters. Suppose we draw a training set of k samples $\mathcal{T} = (s_1, \ldots, s_k)$ independently according to $\mathcal{D}$, and then choose any hypothesis $\sigma_{\mathcal{T}} \in \mathcal{C}$ such that $|\sigma_{\mathcal{T}}(s_i) - \rho(s_i)| \leq \eta$ for all $s_i \in \mathcal{S}$. Then if for $k \geq k_0(\eta, \gamma, \epsilon, \delta)$*

$$\Pr_{s \in \mathcal{S}} \left[|\sigma_{\mathcal{T}}(s) - \rho(s)| > \gamma \right] \leq \varepsilon$$

with probability at least $1 - \delta$ over $\mathcal{T}$, we say that k_0 is a sufficient sample size to learn $\mathcal{C}$.

This says that if the size of the training set, k, is bigger than k_0, then with probability $1-\delta$, the training set $\mathcal{T}$, that we pick according to $\mathcal{D}$ will be a good training set. That is, a hypothesis concept σ which matches the target state on the training set will only be different from the target state on some other sample with small probability, ϵ.

To define a lower bound on k_0, we will need a measure of complexity called the *fat-shattering dimension.*

Definition 7.4 (Aaronson [1])**.** *Let $\mathcal{S}$ be a sample space, let $\mathcal{C}$ be a probabilistic-concept class over $\mathcal{S}$, and let $\gamma > 0$ be a real number. We say a set $\{s_1, \ldots, s_k\} \subseteq \mathcal{S}$ is γ-fat-shattered by $\mathcal{C}$ if there exist real numbers $\alpha_1, \ldots, \alpha_k$ such that for all $B \subseteq \{1, \ldots, k\}$, there exists a probabilistic concept $\rho \in \mathcal{C}$ such that for all $i \in \{1, \ldots, k\}$,*

(i) if $i \notin B$ then $\rho(s_i) \leq \alpha_i - \gamma$, and

(ii) if $i \in B$ then $\rho(s_i) \geq \alpha_i + \gamma$.

Then the γ-fat-shattering dimension of $\mathcal{C}$, or $\mathrm{fat}_{\mathcal{C}}(\gamma)$, is the maximum k such that some $\{s_1, \ldots, s_k\} \subseteq \mathcal{S}$ is γ-fat-shattered by $\mathcal{C}$. (If there is no finite such maximum, then $\mathrm{fat}_{\mathcal{C}}(\gamma) = \infty$.)

The fat-shattering dimension lower bounds the number of samples needed to learn a probabilistic concept.

Lemma 7.5 (Anthony and Bartlett [4])**.** *Suppose $\mathcal{C}$ is a probabilistic concept class over $\mathcal{S}$ and set $0 < \gamma < \eta < 1, \epsilon, \delta \in (0,1)$. Then if $\mathrm{fat}_{\mathcal{C}}(\gamma) \geq d \geq 1$ and $\gamma^2 \geq 4d2^{-\sqrt{d/6}}$, any sample size m_0 sufficient to learn $\mathcal{C}$ satisfies*

$$m_0(\eta, \gamma, \epsilon, \delta) \geq max\left(\frac{1}{32\epsilon}\left(\frac{d}{2\ln^2(4d/\gamma^2)} - 1\right), \frac{1}{\epsilon}\ln\frac{1}{\delta}\right)$$

This concludes the results we will need from statistical learning theory.

7.4 Lower bounds on sample complexity

Our next step is to show that the existence of random access codes lower bounds the fat-shattering dimension. First we have to carefully define what "concept" we will be learning and what constitutes our sample space. For the purposes of learning in GNSTs, the sample space is just the set of possible measurements, where we allow general measurements by first making some fiducial measurement on the state, and then post-processing the result using some decoding function. So we can define $\mathcal{S}_{GNST} := \{(C, D) | C \in \mathcal{E}_G, D : \mathcal{A}^{\times n} \to \{0,1\}\}$. For some sample $(C, D) \in \mathcal{S}_{GNST}$, a concept is specified by the state ρ_x in a GNST via the the probability $\rho_x(C, D) := \sum_{A \in \mathcal{A}^{\times n}} D(A) p_x(A|C)$, where p_x is an n-partite state in some GNST. Then the concept class $\mathcal{C}_{\mathrm{GNST}}$ is the set of concepts specified by all the states in GNST.

Note that a "sample" is stronger than a typical notion of measurement. Usually we say that the measurement gives a result with some probability, but given some sample, the concept ρ actually returns the probability of that outcome occurring. This stronger notion of sampling is all we consider here since we are only lower bounding the number of samples needed.

Claim 7.6. *Let the concept class $\mathcal{C}_{GNST}$ over $\mathcal{S}_{GNST}$ consist of all $\rho_x(C,D) = \sum_{A\in\mathcal{A}} D(A)p_x(A|C)$, where p_x describes any n-partite states in a GNST, over the sample space $\{(C,D)|C \in \mathcal{E}_G, D : \mathcal{A}^n \to \{0,1\}\}$. For integers $n, N(p,n)$ and $\gamma \in (0,1)$, if there exists an $[N(p,n), n, \frac{1}{2}+\gamma]$-RAC then* $\mathrm{fat}_{\mathcal{C}_{GNST}}(\gamma) \geq N$.

Proof. By the RAC definition, there exist a set of measurements $\{(C,D),\ldots,(C^{(N)},D^{(N)})\}$ and states specified by (the concepts) ρ_x for $x \in \{0,1\}^N$ so that

(i) if $x_i = 0$ then $\rho_x(C^{(i)},D^{(i)}) \leq \frac{1}{2} - \gamma$

(ii) if $x_i = 1$ then $\rho_x(C^{(i)},D^{(i)}) \geq \frac{1}{2} + \gamma$

Therefore, this set of samples is γ fat-shattered by $\mathcal{C}_{GNST}$. Since $\mathrm{fat}_{\mathcal{C}_{GNST}}$ is the size of the largest sample set shattered, $\mathrm{fat}_{\mathcal{C}_{GNST}} \geq N(p,n)$. □

Combining Claims 6.5 with 7.6 and 7.5, we get the following result.

Corollary 7.7. *For $\hat{n}$-partite concepts in $\mathcal{C}_{p-GNST}$ and error parameters $\varepsilon,\eta,\gamma,\delta > 0$ with $\gamma > \eta$, if $\hat{n} \geq 2^{2/p}\ln(4/(1-\gamma)^2)$ and*

$$k < max\left(\frac{1}{32\epsilon}\left(\frac{3^{n(\hat{n},p,\gamma)}}{2\ln^2(4\cdot 3^{n(\hat{n},p,\gamma)}/\gamma^2)} - 1\right), \frac{1}{\epsilon}\ln\frac{1}{\delta}\right)$$

for $n(\hat{n},p,\gamma) = \lfloor\left(\frac{\hat{n}\ 2^{-2/p}}{\ln(4/(1-\gamma)^2)}\right)^{\frac{1}{2/p+1}}\rfloor$, then k is not a sufficient sample size to learn states in $\mathcal{C}_{p-GNST}$.

That is, we need $\mathrm{O}(3^{\hat{n}^{\frac{1}{2/p+1}}}/\hat{n}^{\frac{2}{2/p+1}})$ samples to learn an $\hat{n}$-partite state in p-GNST to great accuracy. For $p = 2$ we have an uncertainty relation analogous to quantum mechanics that rules out super-quantum violations of the CHSH bound. Nevertheless it still takes $\mathrm{O}(3^{\sqrt{\hat{n}}}/\hat{n})$ samples to learn these states, as compared to $\mathrm{O}(n)$ in the quantum case.

	p-bin	p-GNST/p-box	p-nonlocal	Quantum	Classical
Non-signaling	yes	yes	yes	yes	yes
Satisfies p-uncertainty	yes	yes	yes	p=2	n/a
Simultaneous measurements	no	local	commuting	commuting	all
CHSH violation	$\frac{1}{2}+\frac{1}{2^{1/p+1}}$	$\frac{1}{2}+\frac{1}{2^{1/p+1}}$	$\frac{1}{2}+\frac{1}{2^{1/p+1}}$	$\frac{1}{2}+\frac{1}{2^{1/2+1}}$	$\frac{3}{4}$
RAC bits to encode N bits	O(polylog(N))	O(polylog(N))	?	$\Omega(N)$	$\Omega(N)$
PIR from N bits	O(polylog(N))	O(polylog(N))	?	$\Omega(N)$	$\Omega(N)$
"Learning" states	hard	hard	?	easy	easy

Table 1: Summary of properties and results for various theories.

8 Consistency of measurements

9 Conclusion and open questions

We have shown that relaxing uncertainty relations can lead to superstrong non-local correlations. This is quite intuitive when considering Tsirelson's bound as a consequence of such an uncertainty relation in the quantum setting. We then constructed a range of theories inspired by such relaxations, and investigated their power with respect to a number of information processing problems. In particular, we obtained superstrong random access encodings and savings for communication complexity. At the same time, however, it turned out to become harder to learn a state in such a theory. We then discussed what makes such superstrong encodings possible in our p-theories, but also in GNST. We identified a number of simple constraints that prevent us from constructing a similar encoding in the quantum setting. Our work may indicate that using "box-world" to understand any other problems within quantum information beyond non-local correlations may be difficult, as "box-world" differs from the quantum setting with respect to such constraints, at least when drawing a one-to-one analogy from a gbit to a qubit as in GNST [8]. It is important to note that these constraints did not prevent us from observing superstrong non-local correlations, but merely forbid our encodings in section 6. If one would like to use "box-world" to understand other aspects one could either impose such consistency constraints, or look for a different approach to defining such theories. GNST was defined by first specifying states and then allowing all operations that take valid states to valid states. If one would have specified the theory in terms of allowed transformations, instead of states, such encodings could also have been ruled out. For example, in the quantum setting one can transform operators $X \otimes X$, $Z \otimes Z$ and $XZ \otimes XZ$ into a bipartite form via a unitary operation. When looking at a density matrix expressed in terms of strings of Pauli matrices, its coefficients (which directly determine the moments for measurements of strings of Paulis) must obey similar constraints to the coefficients belonging to bipartite operators of the form $\mathbb{I} \otimes X, X \otimes \mathbb{I}, X \otimes X$ for example.

Finally, it is clear that both the uncertainty relation and the consistency constraints are obeyed in the quantum setting, since we demand that for any ρ we have $\mathrm{Tr}(\rho) = 1$ and $\rho \geq 0$ to be a valid quantum state. Not surprisingly, both forms of constraints are thus necessary (but in higher dimensions not always sufficient) conditions for $\rho \geq 0$. Such characterizations are not easy for $d > 2$ [26, 10, 21, 41], and it remains an interesting open problem to find an intuitive interpretation for such conditions in higher dimensions, and their consequence for information processing tasks.

10 Acknowledgments

We are indebted to Wim van Dam for useful discussions. This work was supported by NSF grant number PHY-04056720.

References

[1] S. Aaronson. The learnability of quantum states. *Royal Society of London Proceedings Series A*, 463:3089–3114, December 2007.

[2] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower

bound for 1-way quantum automata. In *Proceedings of STOC '99*, pages 376–383, New York, NY, USA, 1999. ACM.

[3] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Quantum dense coding and a lower bound for 1-way quantum finite automata. In *Proceedings of 31st ACM STOC*, pages 376–383, 1999. quant-ph/9804043.

[4] M. Anthony and P. L. Bartlett. Function learning from interpolation. *Comb. Probab. Comput.*, 9(3):213–225, 2000.

[5] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Generalized No-Broadcasting Theorem. *Physical Review Letters*, 99(24):240501–+, December 2007.

[6] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories, 2008.

[7] H. Barnum, O. Dahlsten, M. Leifer, and B. Toner. Nonclassicality without entanglement enables bit commitment. arXiv:0803.1264, 2008.

[8] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304–+, March 2007.

[9] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1965.

[10] R. A. Bertlmann and P. Krammer. Bloch vectors for qudits. *Journal of Physics A: Math. Theor.*, 41:235303, 2008.

[11] G. Brassard, H. Buhrman, N. Linden, A. Methot, A. Tapp, and F. Unger. A limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96:250401, 2006.

[12] S. Bravyi and A. Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A*, 71:022316, 2005.

[13] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter. Implications of superstrong nonlocality for cryptography. *Proceedings of the Royal Society A*, 462(2071):1919–1932, 2006. quant-ph/0504133.

[14] B. Tsirelson (Cirel'son). Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.

[15] B. Tsirelson (Cirel'son). Quantum analogues of Bell inequalities: The case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

[16] J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[17] I. Damgaard, S. Fehr, L. Salvail, and C. Schaffner. Cryptography in the Bounded Quantum-Storage Model. In *Proceedings of 46th IEEE FOCS*, pages 449–458, 2005.

[18] I. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner. A tight high-order entropic quantum uncertainty relation with applications. In *Advances in Cryptology—CRYPTO '07*, volume 4622 of *Lecture Notes in Computer Science*, pages 360–378. Springer-Verlag, 2007.

[19] G. M. D'Ariano. Probabilistic theories: what is special about quantum mechanics?, 2008.

[20] R. de Wolf. Quantum communication and complexity. *Theoretical computer science*, 287(1):337–353, 2002.

[21] K. Dietz. Generalized bloch spheres for m-qubit states. *Journal of Physics A: Math. Gen.*, 36(6):1433–1447, 2006.

[22] A. Doherty, Y.-C. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-provers games. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 199–210, 2008.

[23] M. Forster and S. Wolf. The universality of non-local boxes. In *Proceedings of QCMC*, 2008.

[24] L. Hardy. Quantum Theory From Five Reasonable Axioms. 2001.

[25] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *Proceedings of 35th ACM STOC*, pages 106–115, 2003. quant-ph/0208062.

[26] G. Kimura. The bloch vector for n-level systems. *Physical Review A*, 315:339, 2003.

[27] G. Kimura, T. Miyadera, and H. Imai. Optimal state discrimination in generic probability models, 2008.

[28] L. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Physical Review A*, 73(1):012112–+, January 2006.

[29] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[30] A. Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, 1993.

[31] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[32] S. Popescu and D. Rohrlich. Nonlocality as an axiom for quantum theory. In *The dilemma of Einstein, Podolsky and Rosen, 60 years later: International symposium in honour of Nathan Rosen*, 1996.

[33] S. Popescu and D. Rohrlich. Causality and nonlocality as axioms for quantum mechanics. In *Proceedings of the Symposium of Causality and Locality in Modern Physics and Astronomy: Open Questions and Possible Solutions*, 1997.

[34] M. Seevinck and J. Uffink. Local commutativity versus bell-inequality violation for entangled states and versus non-violation for separable states. *Physical Review A*, 76:042105, 2007.

[35] G. Ver Steeg and S. Wehner. Monogamy of non-local correlations from an uncertainty relation. Unpublished note, 2008.

[36] S.J. Summers. On the independence of local algebras in quantum field theory. *Reviews in Mathematical Physics*, 2(2):201–247, 1990.

[37] B. Toner. Monogamy of nonlocal quantum correlations. quant-ph/0601172, 2006.

[38] B. Toner and F. Verstraete. Monogamy of bell correlations and tsirelson's bound, 2006. quant-ph/0611001.

[39] W. van Dam. Impossible consequences of superstrong nonlocality. quant-ph/0501159, 2005.

[40] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical report, Dept. of Statistics, September 2003.

[41] S. Wehner. Unpublished note. 2008.

[42] S. Wehner and A. Winter. Higher entropic uncertainty relations for anti-commuting observables. *Journal of Mathematical Physics*, 49:062105, 2008.

[43] S. Wehner and J. Wullschleger. Composable security in the bounded-quantum-storage model. In *ICALP 2008*, pages 604–615, 2008.

[44] S. Wolf. Personal communication. 2008.

[45] J. Barrett, N. Linden,S. Massar, S. Pironio, S. Popescu,and D. Roberts. Nonlocal correlations as an information-theoretic resource *Physical Review A*, 71:022101, 2005.